

# aDORe OAI Resource - Overview

## Table of contents

1 What is the aDORe OAI Resource Harvester?.....	2
2 How does the aDORe OAI Resource Harvester work?.....	2
3 Additional Information.....	2

## 1. What is the aDORe OAI Resource Harvester?

The aDORe OAI Resource Harvester is a framework that facilitates collecting resources described by metadata records exposed by OAI-PMH repositories. Metadata records are harvested using the OAI-PMH, and written to write-once/read-many XML wrappers for a collection of XML documents, named XMLTapes. Plug-in based dereferencing implementations that are typically repository-dependent then collect the digital resources described by the harvested records. Collected datastreams are then written to Internet Archive ARC files. The results of these actions are written to control files indicating processing status and the relationship between harvested records and collected digital resources.

## 2. How does the aDORe OAI Resource Harvester work?

- Expose: Producing archives expose OAI Records through an OAI-PMH interface.
- Harvest: Harvested OAI Records are written to XMLTapes.
- Dereference: A plug-in implementation is used to process each OAI Record; downloading resources, generating digests, writing datastreams to ARC files, etc. Since resource processing is typically repository-dependent, the plug-in framework allows developers to create implementations that suit their needs.
- Log: The framework tracks the status of each dereferenced resource; logging success / failure and XMLTape / ArcFile relationships in a Comma-Separated Values file format.

The resulting output of the aDORe OAI Resource Harvester:

- XMLTape: Harvested OAI Records with administrative metadata
- ArcFile: Concatenation of collected datastreams
- ok.csv: Record of each resource successfully harvested to an arc file
- bad.csv: Record of resource harvesting failures

Figure 1

## 3. Additional Information

Bekaert, J., Van de Sompel, H. (2005, June).

[A Standards-based Solution for the Accurate Transfer of Digital Assets](#)

D-Lib Magazine, 11(6)